我在Google search中打入“采药 童子 诗”，AI overview给出“‘采药童子诗’通常指的是唐代诗人贾岛的《寻隐者不遇》...”。搜索结果是正确的，这是我要找的诗。

但有个问题，诗里没说那童子是采药童子，是他老师去采药。为了证明AI没正确理解语义，我打入“西药童子诗”，AI overview给出“‘西药童子诗’实际上是指贾岛的五言古诗《寻隐者不遇》中‘松下问童子，言师采药去’以及‘只在此山中，云深不知处’这两句，并非指代‘西药’的诗歌。” 这就大错特错、驴头不对马嘴了。

我再打入“西药 童子 诗“，结果稍有不同，但仍有错误：“‘西药童子诗’并非一个固定的诗歌名，而是对贾岛《寻隐者不遇》一诗的误传或谐音。” 应该不存在这样的误传或谐音。

所以，计算机语言模型确实不懂语义。还有个问题，在上述三个问题中，AI错了两个，正确率仅1/3。正确率应该是非常不稳定、不可靠的。因为计算机语言模型不懂语义，对每一个正确回答，就可以设计出一大堆相关问题让AI回答错误。

我早就提出统计方法对人工智能测试无效。